Patterns and regulation of dissolved organic carbon: An analysis of 7,500 widely distributed lakes

Sebastian Sobek and Lars J. Tranvik
Limnology / Department of Ecology and Evolution, Uppsala University, Box 573, 75123 Uppsala, Sweden

Yves T. Prairie
Limnology Département des Sciences Biologiques, Université de Québec à Montréal, Case Postale 8888, succ. Centre-ville, Montréal, Québec H3C 3P8, Canada

Pirkko Kortelainen
Finnish Environment Institute, P.O. Box 140, 00251 Helsinki, Finland

Jonathan J. Cole
Institute of Ecosystem Studies, Box AB, Millbrook, New York 12545

Abstract
Dissolved organic carbon (DOC) is a key parameter in lakes that can affect numerous features, including microbial metabolism, light climate, acidity, and primary production. In an attempt to understand the factors that regulate DOC in lakes, we assembled a large database (7,514 lakes from 6 continents) of DOC concentrations and other parameters that characterize the conditions in the lakes, the catchment, the soil, and the climate. DOC concentrations were in the range 0.1–332 mg L\(^{-1}\), and the median was 5.71 mg L\(^{-1}\). A partial least squares regression explained 48% of the variability in lake DOC and showed that altitude, mean annual runoff, and precipitation were negatively correlated with lake DOC, while conductivity, soil carbon density, and soil C:N ratio were positively related with lake DOC. A multiple linear regression using altitude, mean annual runoff, and soil carbon density as predictors explained 40% of the variability in lake DOC. While lake area and drainage ratio (catchment : lake area) were not correlated to lake DOC in the global data set, these two factors explained significant variation of the residuals of the multiple linear regression model in several regional subsets of data. These results suggest a hierarchical regulation of DOC in lakes, where climatic and topographic characteristics set the possible range of DOC concentrations of a certain region, and catchment and lake properties then regulate the DOC concentration in each individual lake.

Dissolved organic carbon (DOC) is a major modulator of the structure and function of lake ecosystems. The DOC pool of lakes consists of both autochthonous DOC (i.e., produced in the lake) and allochthonous DOC (i.e., produced in the catchment), although allochthonous DOC is generally believed to represent the larger fraction of the total DOC in lakes. Due to the dark color of many DOC compounds, DOC affects the thermal structure and mixing depth of lakes (Fee et al. 1996). DOC is an important regulator of ecosystem production, since its absorptive properties impede photosynthesis (Jones 1998), while at the same time, it serves as a substrate for heterotrophic bacteria (Tranvik 1988). These compounded effects of DOC on primary and secondary production result in many lakes being net heterotrophic ecosystems, with the consequence that they export CO\(_2\) to the atmosphere (Sobek et al. 2005). Furthermore, DOC can affect the fate of other solutes, such as metals (Perdue 1998) or organic contaminants (Haitzer et al. 1998). Also, DOC is an efficient protector of lake biota from harmful ultraviolet (UV) radiation (Molot et al. 2004), while hormone-like effects of DOC may affect the physiology of aquatic animals (Steinberg et al. 2004). This wide array of effects on lake ecosystems explains the considerable interest that DOC has received during the last decades.

Principally, the DOC concentration in lakes is regulated in a hierarchical manner (Mulholland et al. 1990). Allochthonous DOC is leached from terrestrial soils to lakes, and the amount of DOC released from soils is determined by the water yield and by the production of leachable organic carbon in soils. In the lakes, autochthonous DOC is produced and, together with allochthonous...
We assembled data on DOC concentrations and adjusted data and the variables with each other. The first component of latent variables, in linear regression and express how much of the variance in the data to variables, and it shows the variance in the data space in order to data. Only latent components is a measure of the predictive power of the PLS model. Variables the degree of explanation than nonlinear regressions, all correlation analyses between variables of the data set were conducted by means of linear regression. Variables with skewed distributions were log-transformed in order to approach normality. The reported \( R^2 \) values were adjusted for the number of parameters in the model. Along with the univariate linear regressions, we also performed multivariate data analysis. We used partial least squares regression (PLS, e.g., Höskuldsson 1988) in order to find out how different variables perform as predictors of lake DOC. The most important advantage of PLS compared to ordinary multiple regression is that PLS is relatively insensitive against dependence of \( X \) variables on each other (co-correlation) and deviations from normality. Also, the PLS algorithm is very tolerant to missing values. A PLS regression extracts for each of the \( X \) variables the degree of explanation of the \( Y \) variable (i.e., DOC), as well as the relationships to the other \( X \) variables, and it shows the results in a two-dimensional scatter plot. In this way, a complex multidimensional data matrix can be dissected and shown in one simple graph. More precisely, PLS regressions place uncorrelated latent components (i.e., orthogonal axes) into the \( X \) data space in order to maximally explain the variance in the \( Y \) data and the \( X \) data. Thereby, PLS uses the variance in the \( X \) data to explain the variance in the \( Y \) data. Only latent components that explain significant variability, as determined by a cross-validation procedure (see Eriksson et al. 2001 for details), are included in the model. The results of a PLS analysis are explored in a series of plots, one of which is the “loadings plot,” which depicts the importance of the different \( X \) variables in explaining the \( Y \) variable, and the relationships of the \( X \) variables with each other. The first component of a PLS (horizontal axis in a loadings plot) always explains more of the variance than the second component (vertical axis), and it is therefore of higher importance when interpreting the loadings plot. The correlation structure of the data set is shown in the spatial distribution of variables in the plot area. Variables are positively correlated when situated close to each other, while they are negatively correlated if they are placed at opposite ends of the plot. Variables in the center of the plot are at most poorly related to the variables at the outer ends of the plot area.

The performance of the PLS model is expressed in the terms \( R^2 Y \), \( R^2 X \), and \( Q^2 \). \( R^2 Y \) and \( R^2 X \) are comparable to \( R^2 \) in linear regression and express how much of the variance in \( Y \) and \( X \) is explained by the latent components. \( Q^2 \) is a measure of the predictive power of the PLS model.

Methods

Data collection—We assembled data on DOC concentrations in the water of global lakes from the published literature, unpublished studies, and national lake surveys. In cases where the concentration of total organic carbon (TOC) was given, it was converted to DOC by multiplying by 0.9 (Wetzel 2001). Along with DOC concentration, we recorded pH and conductivity, as well as lake surface area, catchment area, lake depth, latitude, longitude, altitude, and time of sampling. Since not all of these variables are routinely reported, missing values are frequent in the data set.

To obtain matching soil and climate data, we used ArcGIS 9.0 and fitted the lake data to a number of global-scale databases that are freely available on the internet. Mean annual air temperature and mean annual precipitation were acquired from New et al. (2000), mean annual runoff was acquired from Fekete et al. (2000), and mean annual evaporation and water balance were acquired from Ahn and Tateishi (1994). The density of carbon and nitrogen in the top 1 m of soil was taken from the Global Soil Data Task Group database (2000). These databases consist of gridded data that were extrapolated to the global land area from a network of measurement data. The grid sizes were ~55 km for temperature, precipitation, evaporation, water balance, and runoff, and ~1 km for soil carbon and nitrogen. Data on global land cover, reflecting the terrestrial vegetation or ecosystem type, were acquired from the Global Land Cover Characteristics database (1997), classified according to Olson (1994). Land cover data are based on satellite imagery at a spatial resolution of ~1 km. The lake coordinates were used to match each lake to the database grids within a geographic information system (GIS) framework. For some lakes, matches with the high-resolution databases on soil characteristics and land cover returned missing values and the value “inland water,” respectively. However, many lake coordinates refer to the outlet of the lake rather than the center, and thus only 503 lakes were fitted to the land-cover type “inland water,” independent of lake area.

Statistical analyses—Many of the relationships between the variables in our data set are not strictly linear. However, since linear regression generally yielded higher degrees of explanation than nonlinear regressions, all correlation analyses between variables of the data set were conducted by means of linear regression. Variables with skewed distributions were log-transformed in order to approach normality. The reported \( R^2 \) values were adjusted for the number of parameters in the model. Along with the univariate linear regressions, we also performed multivariate data analysis. We used partial least squares regression (PLS, e.g., Höskuldsson 1988) in order to find out how different variables perform as predictors of lake DOC. The most important advantage of PLS compared to ordinary multiple regression is that PLS is relatively insensitive against dependence of \( X \) variables on each other (co-correlation) and deviations from normality. Also, the PLS algorithm is very tolerant to missing values. A PLS regression extracts for each of the \( X \) variables the degree of explanation of the \( Y \) variable (i.e., DOC), as well as the relationships to the other \( X \) variables, and it shows the results in a two-dimensional scatter plot. In this way, a complex multidimensional data matrix can be dissected and shown in one simple graph. More precisely, PLS regressions place uncorrelated latent components (i.e., orthogonal axes) into the \( X \) data space in order to maximally explain the variance in the \( Y \) data and the \( X \) data. Thereby, PLS uses the variance in the \( X \) data to explain the variance in the \( Y \) data. Only latent components that explain significant variability, as determined by a cross-validation procedure (see Eriksson et al. 2001 for details), are included in the model. The results of a PLS analysis are explored in a series of plots, one of which is the “loadings plot,” which depicts the importance of the different \( X \) variables in explaining the \( Y \) variable, and the relationships of the \( X \) variables with each other. The first component of a PLS (horizontal axis in a loadings plot) always explains more of the variance than the second component (vertical axis), and it is therefore of higher importance when interpreting the loadings plot. The correlation structure of the data set is shown in the spatial distribution of variables in the plot area. Variables are positively correlated when situated close to each other, while they are negatively correlated if they are placed at opposite ends of the plot. Variables in the center of the plot are at most poorly related to the variables at the outer ends of the plot area. The performance of the PLS model is expressed in the terms \( R^2 Y \), \( R^2 X \), and \( Q^2 \). \( R^2 Y \) and \( R^2 X \) are comparable to \( R^2 \) in linear regression and express how much of the variance in \( Y \) and \( X \) is explained by the latent components. \( Q^2 \) is a measure of the predictive power of the PLS model.
If $Q^2$ is similar to $R^2Y$, the model is not overfitted. Calculated models were validated using a permutation test, which results in a measure of the intrinsic “background correlation” of the data, i.e., a baseline in the coefficients $R^2Y$ and $Q^2$ that is not indicative for any correlations between variables but only due to chance. The lower this background correlation, the better and more stable the model. Highly skewed variables (skewness $>2.0$ and min:max ratio $<0.1$) were log-transformed prior to modeling in order to increase model performance; these were lake area, catchment area, drainage ratio, altitude, conductivity, soil C:N, and DOC. All PLS modeling was done on the SIMCA-P 9.0 software (Umetrics AB).

In an attempt to predict lake DOC from a simple model based on a few easily accessible parameters, we performed a multiple linear regression analysis with the $X$ variables that the PLS identified as important predictors of lake DOC. As for the PLS modeling, highly skewed variables were log-transformed prior to analysis in order to approach normality.

**Results**

In total, we managed to compile data from 7,514 lakes, most of which are situated between $25^\circ N$ and $85^\circ N$ (Fig. 1; Table 1). Even though large areas of the world are underrepresented or completely missing (particularly the tropics, Africa, Asia, and South America), the data set covers a wide climatic gradient that stretches from the high Arctic to the subtropics, which allowed us to analyze the relationship between lake DOC concentration and climatic conditions.

Median DOC concentration for all 7,514 lakes was $5.71 \text{ mg L}^{-1}$, ranging from $0.1 \text{ mg L}^{-1}$ to $332 \text{ mg L}^{-1}$. Mean DOC concentration was $7.58 \pm 0.19 \text{ mg L}^{-1}$ ($\pm95\%$ confidence interval). The frequency distribution of DOC was skewed to the right (Fig. 2), and was best modelled with a Gamma distribution:

$$y = \left(\frac{x}{5.886}\right)^{0.970} \times e^{-x/5.886} \times \left[1/5.886 \times \Gamma(1.218)\right]$$  \tag{1}

In $87\%$ of the lakes, DOC was between 1 and $20 \text{ mg L}^{-1}$, while $8.3\%$ of the lakes had concentrations less than $1 \text{ mg L}^{-1}$. Lakes between 20 and $40 \text{ mg L}^{-1}$ were few ($4.2\%$), and only $0.4\%$ of the lakes had DOC concentrations above $40 \text{ mg L}^{-1}$ (Fig. 3A). In $55\%$ of the lakes, DOC concentration was above $5 \text{ mg L}^{-1}$, which has been suggested to be a threshold value for the transition between net autotrophy and net heterotrophy in lakes (Jansson et al. 2000; Prairie et al. 2002). Hence, it is likely that more than half of the lakes in our data set are primarily net heterotrophic ecosystems.

We grouped the lakes according to land cover (Olson 1994), which largely reflects the terrestrial vegetation or ecosystem type (Fig. 3). Grouping the lakes in this way reveals several interesting patterns. First, arctic lakes are generally characterized by low DOC concentrations (Fig. 3B; land-cover types wooded tundra to bare desert, corresponding to an annual mean temperature of $<-4^\circ C$). Second, boreal lakes display a tendency toward higher DOC concentration (Fig. 3B; land-cover types deciduous and mixed boreal forest to deciduous conifer forest, corresponding to an annual mean temperature of $0.5$-
Third, lakes on the northern Great Plains in Saskatchewan, Canada, represent the highest DOC concentrations in the data set (Fig. 3C). We could not distinguish any distinct patterns in DOC for lakes situated in the warmer climate zones (upper part of Fig. 3B).

The PLS regression extracted three significant components from the data set, which collectively explain 49% of the variance in lake DOC ($R^2_Y = 0.48$) and 13% of the variance in the $X$ variables ($R^2_X = 0.13$). The PLS model is highly robust and not overfitted, as model predictability is high ($Q^2 = 0.47$) and background correlation in $R^2_Y$ is low (0.003). The PLS loadings plot (Fig. 4) shows the correlation structure between lake DOC and the different $X$ variables on the first two axes of the model; the third axis explains only 2% of the variability in DOC and is omitted from the plot for reasons of clarity.

The PLS analysis demonstrates that the terrestrial land cover (e.g., conifer boreal forest, barren tundra) strongly affects DOC in lakes (Fig. 4). Interestingly, while the landcover type “conifer boreal forest” is positively related with lake DOC, “cool conifer forest” is negatively related to lake DOC; however, cool conifer forest is confined to high-altitude areas such as the Rocky Mountains and the Alps, which may explain the relatively low DOC concentrations found in these lakes (see discussion of the role of altitude below). Further, the PLS shows that lake DOC is negatively related to mean annual runoff, precipitation, and altitude, and it is positively related to conductivity, soil carbon density, and soil C:N ratio. All these variables explain significant variability in lake DOC. The positive relationship of longitude on lake DOC reflects that many high-DOC lakes are situated in boreal Scandinavia (east of Greenwich, i.e., positive longitude values), and many low-DOC lakes are in the Canadian Arctic and Rocky Mountains (west of Greenwich, i.e., negative longitude values). Neither lake area nor drainage ratio (catchment : lake area) were important predictors of lake DOC concentration (Fig. 4). The relationships between lake DOC and the most important variables of the PLS are also shown as bivariate scatter plots (Fig. 5).
of the variability in lake DOC, indicating that the high discussion of runoff — for soil $R^2$ is 0.58. We found that runoff see also $F_2$ was omitted from the graph for reasons of clarity. The curve depicts the gamma distribution that was fitted to the data (see Results section).

bivariate linear regressions individually explains more than 21% of the variability in lake DOC, indicating that the high predictability of the PLS model ($R^2 = 0.48$) is attributable to the cumulative effect of many variables rather than to the effect of any single variable.

A multiple linear regression, using the most important predictors as identified by the PLS, resulted in the following model (Fig. 5F):

$$\log \text{DOC} = 1.62 + 7.84 \times 10^{-3} \times \text{soil carbon} - 2.96 \times 10^{-4} \times \text{runoff} - 0.41 \times \log \text{altitude};$$

$$R^2 = 0.40; p < 0.0001; F_{3,632} = 1.385$$

where the units are mg L$^{-1}$ for DOC, kg m$^{-2}$ for soil carbon density, mm for mean annual runoff, and meters above sea level for altitude. Including conductivity into the model increased the degree of explanation by 4%; however, conductivity is strongly negatively correlated with altitude ($R^2 = 0.26$) and was therefore excluded from the model. Likewise, precipitation was excluded due to its strong positive correlation with runoff ($R^2 = 0.50$). All other variables in the data set added minimally to the degree of explanation and were therefore not included in the model.

Discussion

The influence of climate and topography on lake DOC—
Climatic conditions are highly important for the DOC concentration in lakes (Figs. 3, 4). In the Arctic, permafrost limits the exportable soil carbon pool, and together with low terrestrial productivity, this results in generally low DOC concentrations in arctic lakes. In the cool climate of boreal forests, soils are often organic-rich because the input of organic matter from the vegetation is high, and the microbial degradation rates are relatively low due to low temperatures and the high frequency of waterlogged soils (e.g., in wetlands). Hence, high-DOC lakes are comparatively frequent in the boreal forest zone. Lakes situated on the northern Great Plains in Saskatchewan frequently lack an outflow, and due to high evaporation, solutes are concentrated in the lake water (Curtis and Adams 1995). Accordingly, the high DOC concentrations in the Saskatchewan lakes are accompanied by high conductivities (up to 74,300 $\mu$S cm$^{-1}$). Hence, climate affects both the production and mobilization of organic carbon from terrestrial soils and the hydrological balance of lakes and therefore plays a superordinate role in the regulation of DOC concentrations in lakes.

The negative relationship between altitude and lake DOC (Fig. 5B; Eq. 2) shows that highland lakes frequently have a lower DOC concentration than lowland lakes. As illustrated by the PLS loadings plot (Fig. 4), the higher in a landscape a lake is situated, the higher the precipitation, which leads to high area-specific runoff. High runoff is in turn associated with low soil carbon density and low lake DOC concentration (Fig. 4; see also discussion of runoff below). Lowland lakes, on the contrary, experience less precipitation and area-specific runoff, which results in longer water residence times and higher soil carbon density. Apart from being related to precipitation and runoff, altitude also carries some information about the landscape topography because catchments can be assumed to become increasingly hilly with increasing altitude. Several studies have shown that catchment slope is negatively correlated with DOC concentration, probably due to thicker organic soil horizons (Rasmussen et al. 1989) and a higher degree of waterlogged soils in catchments with flat topography (D’Arcy and Carignan 1997). Hence, flat topography is coupled to the abundance of wetlands, which in turn is a major predictor of lake DOC both regionally (e.g., Kortelainen 1993) and in lakes from different regions (Xenopoulos et al. 2003). Thus, altitude is a master variable that incorporates climatic, topographic, and edaphic effects on lake DOC.

The effect of runoff on lake DOC—We found that runoff was strongly negatively related to lake DOC concentration (Figs. 4, 5A). This negative correlation was repeated when we divided the data into 73 subsets of lakes from different confined geographical regions (containing at least 15 lakes), and plotted runoff against log DOC for these regional subsets (data not shown). This is surprising because many studies have shown that the export of DOC from terrestrial soils is enhanced by increases in precipitation or runoff (Andersson et al. 1991; Hinton et al. 1997; Mulholland 2003). Likewise, two decades of relative drought were observed to lead to a decline in DOC concentrations (Schindler et al. 1997).

The overall effect of high precipitation and runoff depends on the relationships between the DOC leaching capacity of the soils, the yield and flow path of water, and the fate of DOC in the lake. The DOC leaching capacity of soils at steady state is dependent on the production of leachable organic carbon, and thus it is linked to primary productivity and microbial degradation (Mulholland 2003).
If high runoff and rapid washout of DOC from soils is not balanced by a high production of leachable organic carbon in soils, it will in a long-term perspective lead to a reduced organic carbon export. Further, the flow path of water in the catchment is important. In most soils, a rise in the water table means that the main water flow path gets closer to the soil surface, and the water will therefore be increasingly in contact with the organic-rich uppermost soil horizons (Mulholland 2003), leading to higher DOC export rates. In wetlands, with their water-saturated soils, however, a rising water table will lead to increased surface water flow, and thus a decreased contact of the main water flow path with the organic-rich uppermost soil horizons, leading to lower DOC export (Schiff et al. 1998). Lastly, high runoff results in rapid flushing of lakes, which decreases the opportunity for in-lake loss of DOC (Algesten et al. 2004). On the other hand, high runoff and precipitation lead to an increased input of water to the lake, implying dilution and thus lower concentration of DOC.

The fact that many regional studies on the temporal variability of DOC export have reported positive relationships with runoff may therefore either be due to the fact
that the carbon budget of the studied landscapes is not in steady state (i.e., increased runoff exports more DOC than is produced in soils, implying that observed increases are temporary), or that the changes in runoff are concomitant to changes in leachable organic carbon stocks in soils (i.e., high runoff indicates a high water table, which favors DOC leaching and hampers microbial degradation due to anoxia and humification). Either way, the relationship between lake DOC and runoff presented in this study (Fig. 5A) represents the outcome of the interactions of many climatic and geographic factors over a long time (since the last deglaciation), while the observed changes in DOC export with changes in runoff (Andersson et al. 1991; Hinton et al. 1997; Mulholland 2003) span much shorter time periods (years to decades). Even though the DOC pool available in soils for mobilization and export by water is very large and not easily depleted by fluxes of water (Mulholland 2003), the negative relationship between runoff and lake DOC concentration indicates that when high runoff prevails over extended periods of time, the leachable soil organic carbon pool will eventually be reduced.

The relationship between soil properties and lake DOC—The positive relationship between lake DOC concentration and soil carbon density is in accordance with earlier studies that showed that the export of DOC from catchments is related to the organic carbon stocks in the catchment soils (Hope et al. 1994; Aitkenhead et al. 1999). Hence, we infer that allochthonous DOC is an important contributor to the DOC pool of lakes across a wide climatic gradient (Figs. 4, 5). Since DOC has been recognized as a major modifier of aquatic ecosystems (Wetzel 2001), these results demonstrate that the structure and functioning of lake ecosystems are tightly coupled to the terrestrial environment.

The soil carbon data used in this analysis are a measure of the bulk soil carbon density in the 0–1-m soil profile. Most of the DOC, however, originates from the uppermost organic-rich soil layers, and DOC concentrations in soil pore water decline sharply toward lower soil horizons (McDowell and Likens 1988). We therefore hypothesize that if we had been able to include a measure of organic carbon content in the uppermost soil horizon in our analysis, this variable would have been a more powerful predictor of lake DOC than bulk soil carbon density.

The C:N ratio of the soil explained a small (5%) but significant proportion of the variability in lake DOC (Figs. 4, 5E). It has been shown previously that the soil C:N ratio is a good predictor of the riverine export of
DOC (Aitkenhead and McDowell 2000). The C:N ratio of soils is affected by a multitude of factors (e.g., type and age of vegetation, N deposition, acidification), and a high C:N ratio is indicative of a high proportion of refractory organic matter (e.g., Benner 2003), and thus a lower availability to microbial degradation. In soils with a high C:N ratio, a larger proportion of the soil organic carbon pool may therefore be exported as DOC, rather than respired to CO$_2$ by microbes. Also, as wetland soils often are characterized by a high C:N ratio (Mulholland 2003), the C:N ratio could possibly be a proxy for wetland abundance in the lakes’ surroundings, implying a high export of DOC from the catchment.

Lake area, catchment area, and lake DOC—Lake area is related to lake depth (Wetzel 2001), and, hence, it is connected to lake volume and water retention time and may therefore serve as a proxy for in-lake loss of DOC via mineralization and sedimentation. However, we found that lake area was not an important predictor of lake DOC (Fig. 4), in accordance with a study in lakes from different regions (Xenopoulos et al. 2003). In regional studies, DOC

![Figure 5. Scatter plots and linear regressions for the relationships between log-transformed lake DOC concentration and (A) mean annual runoff ($y = -4.06 \times 10^{-4} \times x + 0.84$; $R^2 = 0.08$; df = 7,075; $p < 0.0001$), (B) log altitude ($y = -0.364 \times x + 1.56$; $R^2 = 0.21$; df = 6,762; $p < 0.0001$), (C) soil carbon density ($y = 9.59 \times 10^{-3} \times x + 0.453$; $R^2 = 0.09$; df = 7,325; $p < 0.0001$), (D) log conductivity ($y = 0.299 \times x + 0.238$; $R^2 = 0.09$; df = 6,695; $p < 0.0001$), (E) log soil C:N ratio ($y = 0.498 \times x + 0.087$; $R^2 = 0.05$; df = 6,501; $p < 0.0001$), and (F) log DOC predicted by the multiple linear regression model $y = 1.618 + 7.84 \times 10^{-3} \times$ soil carbon $-2.96 \times 10^{-4} \times$ runoff $-0.412 \times$ log altitude; $R^2 = 0.40$; df = 6,322; $p < 0.0001$.](image-url)
concentration has been reported to be negatively related to lake area (Rasmussen et al. 1989; Xenopoulos et al. 2003), but no such relationship has been found in other regions (Houle et al. 1995). In Finnish lakes, lake area was not a good predictor of TOC concentration, but nevertheless, median TOC concentration in small lakes was higher than in large lakes (Kortelainen 1993). When dividing our data set into 73 subsets of geographically well-confined regions (containing at least 15 lakes), we found negative relationships between lake area and DOC for several, but not all regions (data not shown). Thus, there was no consistent effect of lake area on the DOC concentration of lakes.

Furthermore, we could not observe an effect of drainage ratio (catchment: lake area) on lake DOC (Fig. 4), again in accordance with the study of lakes in different regions by Xenopoulos et al. (2003). In regional studies, drainage ratio has sometimes been found to be important (Engstrom 1987; Rasmussen et al. 1989; Kortelainen 1993), but sometimes not (Houle et al. 1995). When dividing our data into regional subsets, we found positive relationships between drainage ratio and DOC for a few, but far from all regions (data not shown). Drainage ratio influences lake DOC concentration in two different ways, both of which are in the same direction. First, the DOC loading rate increases with increasing drainage ratio, and second, high drainage ratio implies high flushing rates of lakes and thus low in-lake loss and high concentration of DOC. However, both effects are dependent on water flows and thus climate, which implies that drainage ratio will necessarily be a better predictor of lake DOC within a climatically homogeneous region than across climatic zones. Further, Engstrom (1987) demonstrated that only drainage ratios 4 6–10 will affect lake DOC, which implies that in regions with different topographies, the drainage ratio will be differently important for lake DOC. In our data set, 81% and 64% of the lakes had drainage ratios above 6 and 10, respectively, which, according to Engstrom (1987), would suggest that the DOC concentration in many lakes in our data set is not dependent on drainage ratio.

The observation that drainage ratio and lake area can be important predictors of lake DOC in many regions, but not in lakes distributed across different geographical regions, suggests that other environmental conditions set an average regional DOC level, and that drainage ratio and lake area determine a certain part of the variation around that average regional DOC level in individual lakes. We tested this hypothesis by first using the multiple linear regression model (Fig. 5F; Eq. 2) to predict the log DOC concentration from altitude, runoff, and soil carbon for all lakes. Next, we plotted the residual variation in globally predicted log DOC against log lake area and log drainage ratio for the 73 regional subsets of lakes. There were significant linear relationships (p < 0.05) between the residuals of globally predicted log DOC and log drainage ratio in 26% of the regional subsets (data not shown), explaining on average 15% of the residual variance in log DOC (range 1–40%). Also, for the drainage ratio, the regression statistics were highly variable (ranges for the slope and intercept were −0.33 to 0.54 and −0.93 to 0.22, respectively). Nevertheless, the fact that lake area and drainage ratio explained significant proportions of the residual variance in lake DOC within many geographically confined regions illustrates the hierarchical regulation of lake DOC. The climate, topography, and hydrology set the potential range of lake DOC concentrations within each region, and catchment and lake characteristics (drainage ratio, % wetlands, % upstream lakes, water retention time) regulate the DOC concentration of each individual lake. This implies that regional empirical models on DOC concentration in lakes are limited in their applicability to global extrapolations (Xenopoulos et al. 2003) as well as to other regions, even if situated within relatively short distances (Rantakari et al. 2004).

**Sources of additional variability in lake DOC**—Regional models of organic carbon in lake water rendered degrees of explanation in the range of 31–75% (most frequently 50–70%), and an analysis including lakes from several different regions explained 45% of the variability in lake DOC (Rasmussen et al. 1989; Xenopoulos et al. 2003; Rantakari et al. 2004). Hence, our models compare well to earlier studies. The fact that our analyses leave 52% of the variability in lake DOC unexplained can be attributed to a multitude of factors. First, several important features of the lakes and their catchments were not part of our data set. If we had been able to assemble data on, for example, the proportion of wetlands and upstream lakes and on the volume or flushing rate of the lakes, we probably would have been able to explain a greater share of the variability in lake DOC. Second, the grid size of some of the databases was probably considerably larger than the actual local variability, especially in mountain areas, which adds additional variation to the analyses. Third, most DOC concentrations are derived from one single measurement, which leaves a considerable amount of temporal variability in the data. Fourth, the conversion of TOC to DOC by means of a conversion factor may introduce additional variability in a part of the data set. Last, differences in sampling and analytical protocols among the many studies and surveys will also add to the total unexplained variance.

**Linkages among climate, catchments, and lakes**—This paper shows for the first time that the climatic regulation of DOC concentration in lakes, which has been observed using time series data (Pace and Cole 2002), can also be observed over a wide spatial gradient. Combined with the global-scale importance of the proportion of wetlands in the catchment on lake DOC (Xenopoulos et al. 2003), our study exemplifies the hierarchical structure of the regulation of DOC in lakes: climate and topography regulate terrestrial vegetation, soils, and hydrology, which in turn set the range of possible DOC concentrations in lakes of one region. Then, the DOC concentration of each in-
References


Received: 12 July 2006
Accepted: 1 December 2006
Amended: 30 January 2007